



Jia, Y., Andrieu, C., Piechocki, R.J., & Sandell, M. (2007). Gaussian approximation based mixture reduction for joint channel estimation and detection in MIMO systems. *IEEE Transactions on Wireless Communications*, 6(77), 2384 - 2389.
<https://doi.org/10.1109/TWC.2007.05911>

Peer reviewed version

Link to published version (if available):
[10.1109/TWC.2007.05911](https://doi.org/10.1109/TWC.2007.05911)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Gaussian Approximation Based Mixture Reduction for Joint Channel Estimation and Detection in MIMO Systems

Yugang Jia, *Student Member, IEEE*, Christophe Andrieu, Robert J. Piechocki, *Member, IEEE*, and Magnus Sandell, *Member, IEEE*

Abstract—A novel Gaussian approximation based mixture reduction algorithm is proposed for semi-blind joint channel tracking and symbol detection for spatial multiplexing multiple-input multiple-output (MIMO) systems with frequency-flat time-selective channels. The proposed algorithm is based on a modified sequential Gaussian approximation detector (SGA) [1] which takes into account channel uncertainty, and the first order generalized pseudo-Bayesian (GPB1) channel estimator [2]. Simulation results show that the proposed algorithm performs better than the conventional and computationally expensive decision-directed method with Kalman filter based channel estimation and a posteriori probability (APP) symbol detection.

Index Terms—Joint estimation and detection, MIMO systems, multiple model estimation, multiuser detection, time-varying channels.

I. INTRODUCTION

THE information theoretic results of multiple-input multiple-output (MIMO) systems promise very high data rates with low error probabilities. The performance improvements have also been confirmed in real systems [3] where accurate channel state information (CSI) plays a key role.

A considerable amount of research has been devoted to semi-blind joint channel estimation and detection for MIMO systems with time-selective channels [4], [5]. These methods follow a decision based estimation strategy, i.e. perform symbol detection first and then run a single filter to estimate the channels based on the hard/soft outputs of the symbols. The drawback of these kinds of methods is that the possible symbol detection errors are not fully accounted for in the channel estimation.

An alternative approach is based on the multiple model approach where one operates a bank of Kalman filters for each possible symbol combination and collapses the estimation results from different models into a single Gaussian distribution [6]. However, applying the multiple model algorithm to spatial multiplexing MIMO systems is still computationally

prohibitive for large systems. Suppose the MIMO system has N_T transmit antennas, N_R receive antennas and a modulation symbol alphabet with N symbols: we need to operate N^{N_T} Kalman filters at each time instant where each one has a complexity of $\mathcal{O}((N_T N_R)^3)$. It is however most often the case that the model probabilities of most of the Kalman filters, i.e. the a posteriori probabilities of most symbol combinations, are very small and their outputs contribute very little to the final result of channel estimation as well as symbol detection (marginal posterior probabilities) in each time instant.

In this paper, we propose a novel joint channel estimation and detection algorithm for spatial multiplexing MIMO systems with frequency-flat time-selective channels. Firstly, we modify the sequential Gaussian approximation (SGA) algorithm [1] to take into account channel uncertainty to identify the M most significant symbol combinations. To the best of our knowledge, the sphere decoders (SD) [7] [8] [9], which are also efficient symbol detectors with perfect CSI, have not been extended to take channel uncertainty into consideration for MIMO systems with time-varying channels. We will justify the benefit of considering channel uncertainty in symbol detection in Section V-A. Secondly, we compute channel estimates via a bank of Kalman filters associated with those significant symbol combinations and collapse those estimates via the GPB1 algorithm [2]. To further reduce the complexity, we made the assumption that the rows in channel matrix are independent even when conditioned on the observations (i.e. posterior distribution). Thus, the total complexity of our algorithm is reduced from $\mathcal{O}(M(N_T N_R)^3)$ to $\mathcal{O}(M N_R N_T^2)$ as shown in Section V-B.

II. SYSTEM MODEL

Consider a narrowband spatial multiplexing MIMO system with frequency-flat time-selective channels. At each time instant k , the system model is:

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{n}(k), \quad (1)$$

where $\mathbf{H}(k)$ is the $N_R \times N_T$ Rayleigh flat fading channel matrix with $h_{(i,j)}(k)$ as its (i,j) th entry, which is the channel gain from transmit antenna j to receive antenna i ; $i = 1, \dots, N_R$ and $j = 1, \dots, N_T$; $\mathbf{x}(k) \stackrel{\text{def}}{=} [x_1(k), \dots, x_{N_T}(k)]^T$ ($[\cdot]^T$ means transpose, $[\cdot]^*$ means conjugate and $[\cdot]^H$ means conjugate transpose); a symbol $x_j(k)$ transmitted from the j -th antenna is taken from a modulation constellation $A = \{a_1, a_2, \dots, a_N\}$; $\mathbf{n}(k)$ is a $N_R \times 1$ zero-mean complex circular symmetric Gaussian noise vector with variance matrix $\sigma_n^2 \mathbf{I}$ (\mathbf{I} is the identity matrix).

Manuscript received November 16, 2005; revised March 27, 2006 and August 2, 2006; accepted August 7, 2006. The associate editor coordinating the review of this letter and approving it for publication was P. Martin. This work was supported by Toshiba Research Europe Ltd (Bristol), UK.

Y. Jia was with the University of Bristol, Bristol, UK. He is now with Philips Research East Asia, Shanghai, 200070, P.R. China (e-mail: yugang.jia@philips.com).

R. J. Piechocki is with the Centre for Communications Research, University of Bristol, Bristol, BS8 1UB, UK (e-mail: r.j.piechocki@bristol.ac.uk).

C. Andrieu is with the Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK (e-mail: c.andrieu@bristol.ac.uk).

M. Sandell is with Toshiba Telecommunication Research Lab, Bristol, BS1 4ND, UK (e-mail: magnus.sandell@toshiba-trel.com).

Digital Object Identifier 10.1109/TWC.2007.05911.

The first order AR model (AR1) is widely used for modelling time selective fading channels [5] and will be adopted here for computational reasons (explained in Section V-C):

$$h_{(i,j)}(k) = \alpha h_{(i,j)}(k-1) + v_{(i,j)}(k), \quad (2)$$

where the noise $v_{(i,j)}(k)$ is a zero-mean iid complex circular symmetric Gaussian noise with variance σ_v^2 .

The system described in Eq. (1) and Eq. (2) can be rewritten as follows:

$$\mathcal{H}(k) = \alpha \mathcal{H}(k-1) + \mathbf{v}(k), \quad (3)$$

$$\mathbf{y}(k) = \mathcal{X}(k)\mathcal{H}(k) + \mathbf{n}(k) \quad (4)$$

where $\mathcal{X}(k) \stackrel{\text{def}}{=} \text{Diag}(\mathbf{x}(k)^T, \dots, \mathbf{x}(k)^T)$,

$$\mathcal{H}(k) \stackrel{\text{def}}{=} \begin{bmatrix} h_{(1,1)}(k) \dots h_{(1,N_T)}(k) \\ h_{(2,1)}(k) \dots h_{(2,N_T)}(k) \dots h_{(N_R,N_T)}(k) \end{bmatrix}^T$$

and $\mathbf{v}(k) \stackrel{\text{def}}{=} [v_{(1,1)}(k) \dots v_{(N_R,N_T)}(k)]^T$. We use $\text{Diag}(\cdot)$ for (block) diagonal matrix.

The full a posteriori probability density of the channels for the system described in Eq. (1)-(2) is a Gaussian mixture with the number of components exponential in N_T :

$$p(\mathcal{H}(k)|\mathbf{Y}_{1:k}) = \sum_{\mathbf{x}(1)} \dots \sum_{\mathbf{x}(k)} p(\mathcal{H}(k)|\mathbf{Y}_{1:k}, \mathbf{x}(1), \dots, \mathbf{x}(k)) p(\mathbf{x}(1), \dots, \mathbf{x}(k)|\mathbf{Y}_{1:k}), \quad (5)$$

where $\mathbf{Y}_{1:k} \stackrel{\text{def}}{=} \{\mathbf{y}(1), \dots, \mathbf{y}(k)\}$.

At the k th time instant, we are interested in sequentially updating the a posteriori probabilities of the channel $p(\mathcal{H}(k)|\mathbf{Y}_{1:k})$ as well as the marginal symbol probabilities $p(x_j(k)|\mathbf{Y}_{1:k})$ (for use in channel decoder). This can be achieved using the following:

$$p(\mathcal{H}(k)|\mathbf{Y}_{1:k}) = \sum_{\mathbf{x}(k)} p(\mathcal{H}(k)|\mathbf{Y}_{1:k}, \mathbf{x}(k)) p(\mathbf{x}(k)|\mathbf{Y}_{1:k}), \quad (6)$$

$$p(x_j(k)|\mathbf{Y}_{1:k}) = \sum_{x_1(k)} \dots \sum_{x_{j-1}(k)} \sum_{x_{j+1}(k)} \dots \sum_{x_{N_T}(k)} p(x_1(k), \dots, x_j(k), \dots, x_{N_T}(k)|\mathbf{Y}_{1:k}) \quad (7)$$

where the distribution $p(\mathcal{H}(k)|\mathbf{Y}_{1:k}, \mathbf{x}(k))$ is a Gaussian mixture and the likelihood is given by $p(\mathbf{y}(k)|\mathbf{Y}_{1:k-1}, \mathbf{x}(k))$.

Propagating all the sufficient statistics of the Gaussian distributions is computationally prohibitive. Hence, suboptimal algorithms are sought. In the next section, we introduce strategies to efficiently approximate Eq. (6) and Eq. (7) and outline the main steps of our algorithm.

III. ALGORITHM OUTLINE

A. M best approximation based symbol detection

It is seen that the summations in Eq. (6) and Eq. (7) require a sum of all possible symbol combinations, which is computationally prohibitive for large system. However, the probabilities of most of the symbol combinations are typically very small and contribute very little to the final result. It is possible to replace the summation over all possible symbol combinations with only a subset of M dominant symbol combinations $\Theta_{N_T} \stackrel{\text{def}}{=} \{\mathbf{x}^{(m)}(k) \stackrel{\text{def}}{=} [x_1^{(m)}(k), \dots, x_{N_T}^{(m)}(k)]^T, m =$

$1, \dots, M\}$ as in [1]. This will result in the following approximation:

Approximation 1: The summation of N^{N_T} combinations can be approximated with that of the M dominant ones:

$$p(\mathcal{H}(k)|\mathbf{Y}_{1:k}) \approx \sum_m p(\mathcal{H}(k)|\mathbf{Y}_{1:k}, \mathbf{x}^{(m)}(k)) p(\mathbf{x}^{(m)}(k)|\mathbf{Y}_{1:k}), \quad (8)$$

$$\begin{aligned} p(x_j(k)|\mathbf{Y}_{1:k}) &\approx \sum_m p\left(x_1^{(m)}(k), \dots, x_j^{(m)}(k), \dots, x_{N_T}^{(m)}(k)|\mathbf{Y}_{1:k}\right) \\ &\times \mathbb{I}\left(x_j^{(m)}(k) = x_j(k)\right) \end{aligned} \quad (9)$$

where $\mathbb{I}\left(x_j^{(m)}(k) = x_j(k)\right)$ is the indicator function for the event $\left(x_j^{(m)}(k) = x_j(k)\right)$.

We will present a suboptimal Gaussian approximation based mixture reduction method to find the M dominant symbol combinations in Section IV-A. Then this suboptimal identification procedure and Approximation 1 (Eq. (9)) will be justified via computer simulations in Section V-A.

B. Reduced Complexity Kalman Filter Based Channel Estimator

The update of $p(\mathcal{H}(k)|\mathbf{Y}_{1:k}, \mathbf{x}^{(m)}(k))$ in Eq. (8) requires operating Kalman filters associated with a specific symbol combination $\mathbf{x}^{(m)}(k)$ which has a complexity of $(N_T N_R)^3$. To further reduce the complexity, we propose the following approximation:

Approximation 2: The joint distribution of the rows in $\mathbf{H}(k)$ can be approximated with the product of the marginal distributions of each row:

$$p(\mathbf{H}_{(1,:)}(k), \dots, \mathbf{H}_{(N_R,:)}(k)|\mathbf{Y}_{1:k}) \approx \prod_i p(\mathbf{H}_{(i,:)}(k)|\mathbf{Y}_{1:k}) \quad (10)$$

where $\mathbf{H}_{(i,:)}(*)$ is the i th row of $\mathbf{H}(*)$.

Hence we focus on the approximation of the marginal probabilities $p(\mathbf{H}_{(i,:)}(k)|\mathbf{Y}_{1:k})$ that is, more specifically, the approximation of the two moments $\hat{\mathbf{H}}_{(i,:)}(k|k) \stackrel{\text{def}}{=} E(\mathbf{H}_{(i,:)}(k)|\mathbf{Y}_{1:k})$ and $\mathbf{P}_i(k|k) \stackrel{\text{def}}{=} E\left(\left(\mathbf{H}_{(i,:)}(k) - \hat{\mathbf{H}}_{(i,:)}(k|k)\right)^T \left(\mathbf{H}_{(i,:)}(k) - \hat{\mathbf{H}}_{(i,:)}(k|k)\right)^* | \mathbf{Y}_{1:k}\right)$ for $i = N_1, \dots, N_T$. Using this approximation, the complexity of the Kalman filter is dramatically reduced to $\mathcal{O}(N_T N_R^2)$ with only slight performance degradation as shown in Sections V-A and V-B.

The final channel estimates at the k -th time instant is obtained via the GPB1 algorithm which collapses the channel estimates from M Kalman filters, which will be described in Section IV-B. A general introduction of the GPB1 (multiple model) algorithm can be found in [2] [10] and a detailed description of our system can also be found in [11].

IV. ALGORITHM DESCRIPTION

We first start this section with some notations related to $p(\mathcal{H}(k))$, and in particular introduce a representation of $\mathcal{H}(k)$

used in the sequel. Suppose that at time $k-1$, the distribution of interest $p(\mathcal{H}(k-1)|\mathbf{Y}_{1:k-1})$ can be approximated as a complex circular symmetric Gaussian distribution with mean $\hat{\mathcal{H}}(k-1|k-1)$ and covariance $\mathcal{P}(k-1|k-1)$ defined as follows:

$$\hat{\mathcal{H}}(k-1|k-1) \stackrel{\text{def}}{=} E(\mathcal{H}(k-1)|\mathbf{Y}_{1:k-1}),$$

$$\mathcal{P}(k-1|k-1) \stackrel{\text{def}}{=} E\left(\left(\hat{\mathcal{H}}(k-1|k-1) - \mathcal{H}(k-1)\right) \times \left(\hat{\mathcal{H}}(k-1|k-1) - \mathcal{H}(k-1)\right)^H | \mathbf{Y}_{1:k-1}\right).$$

As a result of Approximation 2, the variance matrix $\mathcal{P}(k-1|k-1) = \text{Diag}(\mathbf{P}_1(k-1|k-1), \dots, \mathbf{P}_{N_R}(k-1|k-1))$ is a block diagonal matrix.

With the Kalman predictor for AR1 model in Eq. (2), we can obtain channel prediction $\hat{\mathcal{H}}(k|k-1) = \alpha \hat{\mathcal{H}}(k-1|k-1)$ and prediction variance $\mathcal{P}(k|k-1) = \alpha^2 \mathcal{P}(k-1|k-1) + \sigma_v^2 \mathbf{I}$ with $\hat{\mathbf{H}}(*) = [\hat{\mathbf{H}}_{(1,:)}^T(*) \dots \hat{\mathbf{H}}_{(N_R,:)}^T(*)]$ and $\tilde{\mathcal{H}}(*) = [\hat{\mathbf{H}}_{(1,1)}(*) \dots \hat{\mathbf{H}}_{(1,N_T)}(*) \dots \hat{\mathbf{H}}_{(N_R,1)}(*) \dots \hat{\mathbf{H}}_{(N_R,N_T)}(*)]^T$. Hence one can represent $\mathcal{H}(k) = \tilde{\mathcal{H}}(k|k-1) + \tilde{\mathcal{H}}(k|k-1)$ where $\tilde{\mathcal{H}}(k|k-1)$ is a zero mean circular symmetric Gaussian random variable with variance $\mathcal{P}(k|k-1)$ and rewrite Eq. (1) as follows:

$$\mathbf{y}(k) = \hat{\mathbf{H}}(k|k-1)\mathbf{x}(k) + \mathcal{X}(k)\tilde{\mathcal{H}}(k|k-1) + \mathbf{n}(k). \quad (11)$$

A. *M* Best Significant Symbol Combinations Identification and Marginal Symbol Probabilities Computation

First, we explain how to identify the M most significant symbol combinations via a modified SGA algorithm [1] with channel uncertainty. Given that we have identified M significant combinations $\Theta_{j-1}(k) \stackrel{\text{def}}{=} \{x_1^{(m)}(k), \dots, x_{j-1}^{(m)}(k), m = 1, 2, \dots, M\}$ for antenna $1, 2, \dots, j-1$ at the $(j-1)$ -th step of the algorithm. We would like to calculate

$$\begin{aligned} & p(x_1^{(m)}(k), \dots, x_{j-1}^{(m)}(k), x_j(k) | \mathbf{Y}_{1:k}) \\ & \propto p(\mathbf{y}(k) | x_1^{(m)}(k), \dots, x_{j-1}^{(m)}(k), x_j(k), \mathbf{Y}_{1:k-1}) \\ & \times p(x_j(k)) \prod_{l=1}^{j-1} p(x_l^{(m)}(k)) \\ & \stackrel{\text{def}}{=} \psi_m(x_j(k)) \end{aligned}$$

for all $m = 1, \dots, M$ and $x_j(k) \in A$ in order to select $\Theta_j(k)$ which contains M symbol combinations of the largest probabilities, among the MN possibilities.

However, the computation of MN likelihoods is still prohibitive for large systems. We can rewrite Eq. (11) as follows,

$$\begin{aligned} \mathbf{y}(k) &= \sum_{l=1}^j \hat{\mathbf{H}}_{(:,l)}(k|k-1)x_l(k) \\ &+ \sum_{l=j+1}^{N_T} \hat{\mathbf{H}}_{(:,l)}(k|k-1)x_l(k) \\ &+ \mathcal{X}(k)\tilde{\mathcal{H}}(k|k-1) + \mathbf{n}(k) \\ &\stackrel{\text{def}}{=} \sum_{l=1}^j \hat{\mathbf{H}}_{(:,l)}(k|k-1)x_l(k) + \tilde{\mathbf{n}}_j(k). \quad (12) \end{aligned}$$

where $\mathbf{H}_{(:,l)}(*)$ is the l th column of $\mathbf{H}(*)$ for $l = 1, \dots, N_T$.

Here we approximate the interference noise term $\tilde{\mathbf{n}}_j(k)$ as a moment matched Gaussian distribution which is known as probabilistic data association (PDA) in the literature [12] [13] [14] [15] [16]. We can calculate an approximation to $\psi_m(x_j(k))$ as follows:

$$\begin{aligned} \psi_m(x_j(k)) &\approx \exp\left(-\left(\mathbf{w}_j^{(m)}(k)\right)^H \mathbf{\Pi}_j^{-1}(k) \mathbf{w}_j^{(m)}(k)\right) \\ &\times p(x_j(k)) \prod_{l=1}^{j-1} p(x_l^{(m)}(k)) \quad (13) \end{aligned}$$

$$\begin{aligned} \mathbf{w}_j^{(m)}(k) &= \mathbf{y}(k) - \sum_{l=1}^{j-1} \hat{\mathbf{H}}_{(:,l)}(k|k-1)x_l^{(m)}(k) \\ &\quad - \hat{\mathbf{H}}_{(:,j)}(k|k-1)x_j(k), \end{aligned}$$

$$\begin{aligned} & \mathbf{\Pi}_j^{-1}(k) \\ &= \left(\mathbf{\Pi}(k) + \gamma \sum_{l=j+1}^{N_T} \hat{\mathbf{H}}_{(:,l)}(k|k-1) \left(\hat{\mathbf{H}}_{(:,l)}(k|k-1) \right)^H \right)^{-1}, \end{aligned}$$

$$\mathbf{\Pi}(k) = \sigma_n^2 \mathbf{I} + \text{Var}\left(\mathcal{X}(k)\tilde{\mathcal{H}}(k|k-1)\right),$$

$$\begin{aligned} & \text{Var}\left(\mathcal{X}(k)\tilde{\mathcal{H}}(k|k-1)\right) \\ &= \gamma \text{Diag}(\text{Tr}(\mathbf{P}_1(k|k-1)), \dots, \text{Tr}(\mathbf{P}_{N_R}(k|k-1))), \end{aligned}$$

where the mean of the modulation alphabet A is zero and its variance is γ (w.r.t. a uniform distribution) and $\text{Tr}(\cdot)$ means the trace of a matrix. The matrix $\mathbf{\Pi}_j^{-1}(k)$ for $j = 1, \dots, N_T-1$ can be computed sequentially via the matrix inversion lemma given in [11].

Then M symbol combinations with the largest $\psi_m(x_j(k))$ are selected among the MN possible symbol combinations, resulting in a new set $\Theta_j(k)$. At the end of this selection process, we can obtain the set $\Theta_{N_T}(k)$ which contains M of the most significant symbol combinations $\mathbf{x}^{(m)}(k)$, $m = 1, \dots, M$.

Then, the marginal symbol probabilities can be computed from the MN likelihoods $\psi_m(x_{N_T}(k))$, $m = 1, \dots, M$ and $x_{N_T}(k) \in A$ with approximation 1 (Eq.(9)). For $j = 1, \dots, N_T-1$:

$$\begin{aligned} p(x_j(k) | \mathbf{Y}_{1:k}) &\approx \frac{1}{\mathcal{Z}(k)} \sum_{x_{N_T}(k) \in A} \sum_m \psi_m(x_{N_T}(k)) \\ &\times \mathbb{I}\left(x_j^{(m)}(k) = x_j(k)\right) \quad (14) \end{aligned}$$

where $\mathcal{Z}(k)$ is a normalizing constant. For the N_T th antenna, $p(x_{N_T}(k) | \mathbf{Y}_{1:k}) \approx \frac{1}{\mathcal{Z}(k)} \sum_m \psi_m(x_{N_T}(k))$.

B. GPB1 Channel Estimation

Operating a Kalman filter for each identified significant symbol combination $\mathbf{x}^{(m)}$, $m = 1, \dots, M$, we can get channel estimation $\hat{\mathbf{H}}_{(i,:)}^{(m)}(k|k)$, estimation variance $\mathbf{P}_i^{(m)}(k|k)$, measurement residual $\mathbf{v}_i^{(m)}(k)$ and residual variance $\Omega_i^{(m)}(k)$ for $i = 1, \dots, N_T$ as described in [11].

Then the model probabilities can be computed as follows:

$$p(\mathbf{x}^{(m)}(k)|\mathbf{Y}_{1:k}) = \phi_m(k) / \sum_m \phi_m(k), \quad (15)$$

$$\begin{aligned} \phi_m(k) &= \exp\left(-\sum_i \left(|\mathbf{v}_i^{(m)}(k)|^2 / \Omega_i^{(m)}(k)\right)\right) \\ &\times \prod_j p(x_j^{(m)}(k)) / \prod_i \Omega_i^{(m)}(k). \end{aligned}$$

Collapsing the channel estimates via the GPB1 algorithm, we get $\mathcal{H}(k|k)$ for the next time instant:

$$\begin{aligned} \hat{\mathbf{H}}_{(i,:)}(k|k) &= \sum_{\mathbf{x}(k)} E(\mathbf{H}_{(i,:)}(k)|\mathbf{Y}_{1:k}, \mathbf{x}(k)) \\ &\times p(\mathbf{x}(k)|\mathbf{Y}_{1:k}) \\ &\approx \sum_m \hat{\mathbf{H}}_{(i,:)}^{(m)}(k|k) p(\mathbf{x}^{(m)}(k)|\mathbf{Y}_{1:k}). \end{aligned}$$

The co-variance $\mathcal{P}(k|k)$ for the next time instant is shown in top of the next page where the underlined term is known as the “spread-of-the-means” term [2][11].

V. SIMULATION RESULTS

In this section, we provide computer simulation examples to compare the performance of the proposed SGAGPB algorithm ($M = 20$) with that of the APP detector with known CSI for each time instant (APPKnowChan), the genie aided method (performance bound) and the baseline system which used the APP / SD [9] detectors and a Kalman filter channel estimation based on soft symbol decisions (SDKal/APPKal). The channel uncertainty is not included in symbol detection in the SDKal/APPKal algorithms. In the genie-aided approach, we estimate the channels using a randomly generated symbol sequence known to the receiver via a Kalman filter and then detect the symbols via the APP detector with channel estimation. In order to validate the approximations, we also provide simulation results for the modified SGA algorithm ($M = 20$) with a Kalman filter based channel estimator (SGAKal) and GPB1 based channel estimator with full covariance (SGAGPB-FullCov) respectively. The SGAGPB-FullCov algorithm works without approximation 2 but with approximation 1.

The simulation is based on a coded MIMO system with soft non-iterative decoding and detection (i.e. the MIMO detector processes the data only once) [17]. We set $N_T = N_R = 4$ and consider a 16QAM modulation with 1152 bits per frame before channel coding. A 1/2 rate Turbo Coder with generators 7 and 5 in octal notation is used at the transmitter and a BCJR channel decoder with 4 iterations is used at the receiver. For each block, the Rayleigh fading channel is generated by Clarke’s model [18] with a normalized maximum Doppler spread $f_d = 1e - 3$. The channels related to each transmitter and receiver pair are generated independently. The SNR is defined as $E\{||\mathbf{H}\mathbf{x}||^2\} / E\{||\mathbf{n}||^2\} = \gamma N_T / \sigma_n^2$.

The initial channel estimation $\hat{\mathbf{H}}(0|0)$ for all the algorithms is computed from the training sequence via maximum likelihood estimation [19] $\hat{\mathbf{H}}(0|0) = \mathbf{Y}(0)\mathbf{X}(0)^H$ where $\mathbf{X}(0)$ is a $N_T \times N_T$ orthogonal training sequence known to the receiver

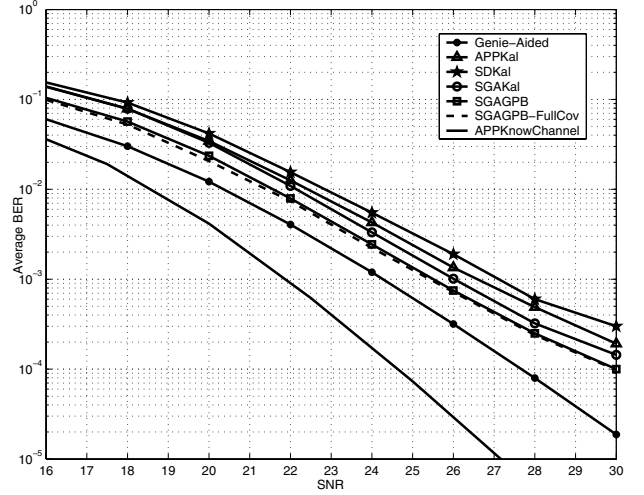


Fig. 1. Uncoded BER performance of various algorithms for a 16QAM, 4×4 MIMO system with $f_d = 10^{-3}$.

and $\mathbf{Y}(0)$ is the observation matrix at receiver. The initial channel estimation variance is $\mathcal{P}(0|0) = 0.002\mathbf{I}$ and $\alpha \approx 1$.

It is possible that some of the marginal symbol probabilities calculated from the SGA based algorithms via Approximation 1 in Eq. (9) will be zero. Thus a limit is set for the soft output (Log likelihood ratio (LLR) = $[-20, 20]$) of the SGA based algorithms to get the best performance.

A. Performance Comparison

Fig. 1 illustrates the uncoded BER performance of all the algorithms. It can be seen that the SGAGPB algorithm is 1 dB better than the APPKal algorithm and just 1 dB worse than the genie-aided approach (performance bound) in the uncoded case. Fig. 2 shows the channel estimation mean square error (MSE) for different algorithms. The better channel estimation provided by the SGAGPB algorithm results in good soft decoding quality as shown in the coded BER performance in Fig. 3. Simulation results for other MIMO systems (e.g. $N_T = N_R = 6$ or turbo receiver systems) can be found in [11].

B. Complexity Comparison

The Kalman filter used in the APPKal, SDKal, SGAKal and SGAGPB algorithms has a total complexity of $\mathcal{O}(N_R N_T^2)$ (the covariance matrix $\mathcal{P}(k-1|k-1)$ is block diagonal). The complexity of computing the $M N N_T$ likelihoods in Eq. (13) is $\mathcal{O}(M N N_T N_R^2)$ complex operations. Therefore, the overall complexity of the SGAGPB algorithm is approximately $\mathcal{O}(M N N_T N_R^2)$ complex operations. Fig. 4 summarizes the total number of operations (real ADD+MUL) of the APPKal, SDKal, SGAKal, SGAGPB and SGAGPB-FullCov algorithms for each time instant. The number of operations of the SD algorithm is averaged from 1000 channel realizations with SNR=22 dB. The complexity of the SGAGPB and SGAKal algorithms increases far slower than that of the APPKal and SDKal algorithms with increasing antenna numbers.

$$\begin{aligned}
\mathbf{P}_i(k|k) &= \sum_{\mathbf{x}(k)} E \left(\left(\mathbf{H}_{(i,:)}(k) - \hat{\mathbf{H}}_{(i,:)}(k|k) \right)^T \left(\mathbf{H}_{(i,:)}(k) - \hat{\mathbf{H}}_{(i,:)}(k|k) \right)^* | \mathbf{Y}_{1:k}, \mathbf{x}(k) \right) p(\mathbf{x}(k) | \mathbf{Y}_{1:k}) \\
&\approx \sum_m \left(\left(\hat{\mathbf{H}}_{(i,:)}^{(m)}(k|k) - \hat{\mathbf{H}}_{(i,:)}(k|k) \right)^T \left(\hat{\mathbf{H}}_{(i,:)}^{(m)}(k|k) - \hat{\mathbf{H}}_{(i,:)}(k|k) \right)^* + \mathbf{P}_i^{(m)}(k|k) \right) p(\mathbf{x}^{(m)}(k) | \mathbf{Y}_{1:k}).
\end{aligned}$$

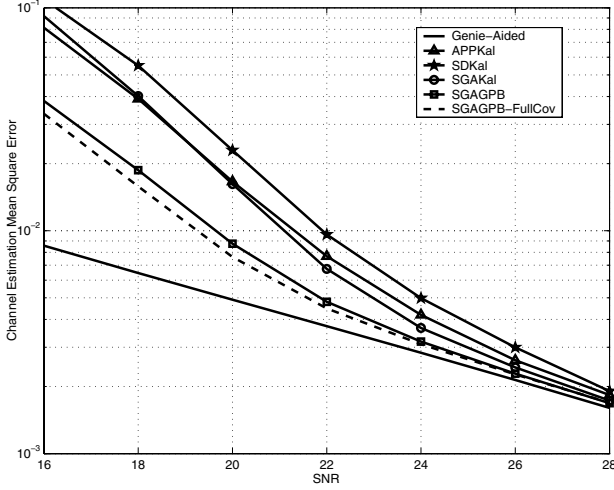


Fig. 2. Channel estimation MSE of various algorithms for a 16QAM, 4×4 MIMO system with $f_d = 10^{-3}$.

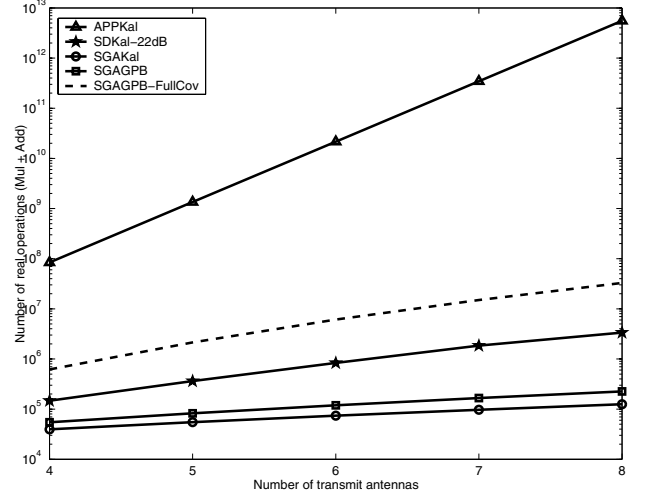


Fig. 4. Complexity comparison of various algorithms for a 16QAM, MIMO system with different values of N_T .

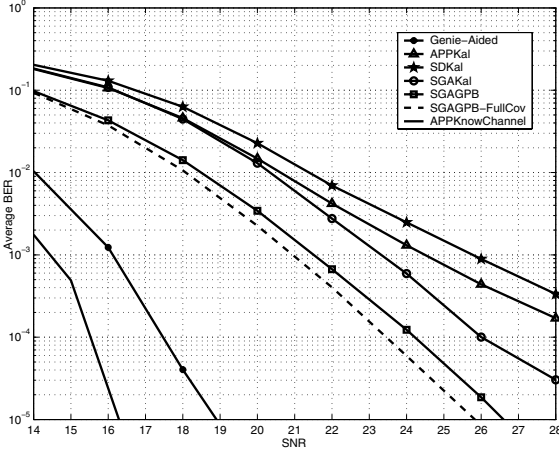


Fig. 3. Coded BER performance of various algorithms for a 16QAM, 4×4 MIMO system with $f_d = 10^{-3}$.

C. Discussion

The performance of the SGAKal algorithm is better than that of the SDKal and APPKal algorithms. This justifies the proposed identification procedure and marginal symbol probabilities computation method (Eq. (9) in Approximation 1) and shows that it is beneficial to take into account channel uncertainty for symbol detection.

The SGAGPB algorithm outperforms the SGAKal algorithm. This demonstrates that the GPB1 based multiple model channel estimator works better than a Kalman filter channel estimator at slightly increased complexity (Eq. (8) in Approximation 1). The complexity of the SGAGPB algorithm

with Approximation 2 (SGAGPB) is much lower than that of the SGAGPB-FullCov algorithm with slightly degraded performance. This shows that there is a good tradeoff between complexity and performance when using Approximation 2.

The proposed algorithm is based on the GPB1 algorithm and only works for first order AR channel models. For $l > 1$ -th order AR channel model, it is necessary to consider l -th order GPB algorithm and joint symbol detection of l time instants which is not trivial for implementation. An efficient suboptimal solution is still an open problem.

VI. CONCLUSIONS

We have proposed a new joint channel estimation and symbol detection scheme for MIMO systems based on the Gaussian approximation and the GPB1 algorithm. First, we have modified the SGA algorithm for identification of the M most significant symbol combinations in order to take into account channel uncertainty. Then reduced complexity Kalman filtering is performed for each symbol combination and all the estimations from different models are collapsed into a final one, which is propagated to the next time instant. Simulation results show that the performance of the proposed algorithm is much better than that of the APP detector with single Kalman filter based channel estimation algorithm while enjoying lower complexity.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the fruitful discussions with the researchers at Toshiba Telecommunication Research Lab (Bristol) and the support of its directors. They also

would like to thank anonymous reviewers for their constructive comments and Mr. C. M. Vithanage and Mr. M. Webb for their great help with grammar.

REFERENCES

- [1] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian approximation based mixture reduction for near optimal detection for MIMO system," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 997–999, Nov. 2005.
- [2] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques and Software*. Dedham, MA: Artech House, 1993.
- [3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.
- [4] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [5] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time selective fading channels," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 183–186, Feb. 2002.
- [6] H. Kulatunga and V. Kadiramanathan, "Adaptive joint detection and estimation in MIMO systems: A hybrid systems approach," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1629–1644, May 2006.
- [7] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short lengths in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, no. 3, pp. 463–471, Apr. 1985.
- [8] D. Pham, J. Luo, K. Pattipati, and P. Willett, "An improved complex sphere decoder for V-BLAST systems," *IEEE Signal Processing Lett.*, vol. 9, no. 11, pp. 748–751, May 2004.
- [9] M. S. Yee, "Max-log-map sphere decoder," in *Proc. IEEE ICASSP*, Mar. 2005, pp. 1013–1016.
- [10] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part V: Multiple-model methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1255–1321, Oct. 2005.
- [11] Y. Jia, C. Andrieu, R. J. Piechocki, and M. Sandell, "Gaussian approximation based mixture reduction for joint channel estimation and detection in MIMO systems," Centre for Communications Research, University of Bristol, Tech. Rep., 2006. [Online]. Available: <http://www2.maths.bris.ac.uk/~eexyj/JIA06SGAGPBTR>
- [12] J. Luo, K. Pattipati, P. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, Sep. 2001.
- [13] D. Pham, K. Pattipati, P. Willett, and J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 205–207, Apr. 2004.
- [14] S. Liu and Z. Tian, "Near-optimum soft decision equalization for frequency selective MIMO channels," *IEEE Trans. Signal Processing*, vol. 5, no. 3, pp. 721–733, Mar. 2004.
- [15] J. C. Fricke, M. Sandell, J. Mietzner, and P. A. Hoehner, "Impact of the Gaussian approximation on the performance of the probabilistic data association MIMO decoder," *EURASIP J. Wireless Communications Networking*, vol. 2005, no. 5, pp. 796–800, Oct. 2005.
- [16] Y. Jia, C. M. Vithanage, C. Andrieu, and R. J. Piechocki, "Probabilistic data association for symbol detection in MIMO systems," *IEEE Electron. Lett.*, vol. 42, no. 1, pp. 38–40, Jan. 2006.
- [17] B. M. Hochwald and S. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
- [18] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [19] B. Hassibi and B. Hochwald, "How much training is needed in a multiple-antenna wireless link?" *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 951–964, Apr. 2003.